

T S1/5/1

1/5/1

DIALOG(R) File 351:Derwent WPI

(c) 2005 Thomson Derwent. All rts. reserv.

012697796 **Image available**

WPI Acc No: 1999-503905/199942

XRPX Acc No: N99-376716

Duplicate data co-ordination management procedure in distributed database system - involves performing strong and weak co-ordination management at different levels depending on variety of transaction occurred in system

Patent Assignee: JISEDAL JOHO HOSO SYSTEM KENYUJO KK (JISE-N); RICOH KK (RICO)

Number of Countries: 001 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 11219309	A	19990810	JP 98112305	A	19980422	199942 B
JP 3563591	B2	20040908	JP 98112305	A	19980422	200459

Priority Applications (No Type Date): JP 97328217 A 19971128; JP 97264643 A 19970929

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
JP 11219309	A		21	G06F-012/00	
JP 3563591	B2		29	G06F-012/00	Previous Publ. patent JP 11219309

Abstract (Basic): JP 11219309 A

NOVELTY - Strict co-ordination management is done enabling immediate and delayed updating transmission from main site to other sites respectively for first two levels. Weak co-ordination management is done for third level, delaying updating transmission from main site to other site. DETAILED DESCRIPTION - During transaction, the level of co-ordination control for reading or updating duplicate data is decided based on transaction variety. An INDEPENDENT CLAIMS are also included for the following: duplicate data co-ordination management procedure; recording medium stored with computer readable co-ordination management program

USE - In distributed database system.

ADVANTAGE - Optimize co-ordination control as it is done depending on variety of transaction.

Dwg.1/14

Title Terms: DUPLICATE; DATA; CO; ORDINATE; MANAGEMENT; PROCEDURE; DISTRIBUTE; DATABASE; SYSTEM; PERFORMANCE; STRONG; WEAK; CO; ORDINATE; MANAGEMENT; LEVEL; DEPEND; VARIETY; TRANSACTION; OCCUR; SYSTEM

Derwent Class: T01

International Patent Class (Main): G06F-012/00

International Patent Class (Additional): G06F-015/16; G06F-017/30

File Segment: EPI

?

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-219309

(43) 公開日 平成11年(1999) 8月10日

(51) Int. Cl. ⁶	識別記号	F I			
G06F 12/00	518	G06F 12/00	518	A	
	533		533	J	
15/16	370	15/16	370	M	
17/30		15/40	310	C	
			310	F	

審査請求 未請求 請求項の数 7 O L (全21頁) 最終頁に続く

(21) 出願番号	特願平10-112305	(71) 出願人	000006747 株式会社リコー 東京都大田区中馬込 1 丁目 3 番 6 号
(22) 出願日	平成10年(1998) 4月22日	(71) 出願人	597136766 株式会社次世代情報放送システム研究所 東京都台東区西浅草 1 丁目 1-1
(31) 優先権主張番号	特願平9-264643	(72) 発明者	吉浦 由香利 東京都大田区中馬込 1 丁目 3 番 6 号 株式 会社リコー内
(32) 優先日	平 9 (1997) 9月29日	(72) 発明者	飯沢 篤志 東京都大田区中馬込 1 丁目 3 番 6 号 株式 会社リコー内
(33) 優先権主張国	日本 (J P)	(74) 代理人	弁理士 酒井 宏明
(31) 優先権主張番号	特願平9-328217		
(32) 優先日	平 9 (1997) 11月28日		
(33) 優先権主張国	日本 (J P)		

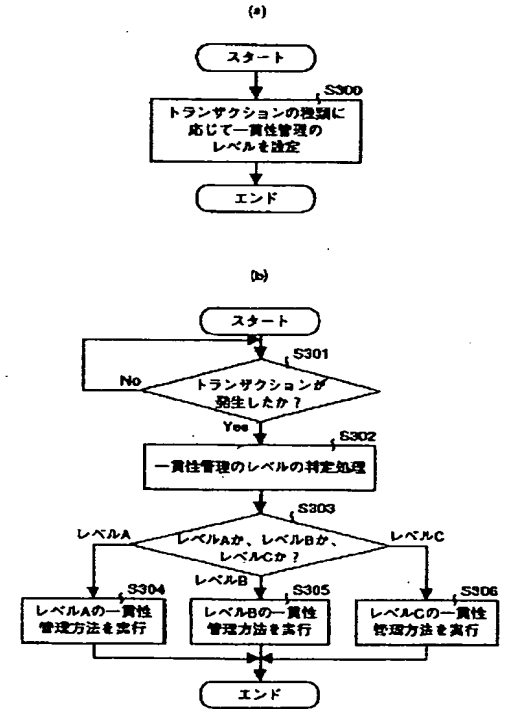
最終頁に続く

(54) 【発明の名称】 分散型データベースシステムの一貫性管理方法およびその方法の各工程をコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体

(57) 【要約】

【課題】 トランザクションの種類に応じて異なる一貫性管理を行うこと。

【解決手段】 各トランザクションの種類に一貫性管理のレベルを表す第1レベル、第2レベルおよび第3レベルのいずれかを設定し (S 3 0 0)、システム内においてトランザクションが発生した場合に (S 3 0 1)、発生したトランザクションの種類に基づいて一貫性管理のレベルを判定し (S 3 0 2、S 3 0 3)、第1レベルの場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、主サイトから各サイトへの更新伝播を即時行い (S 3 0 4)、第2レベルの場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、主サイトから各サイトへの更新伝播を遅延させて行い (S 3 0 5)、第3レベルの場合に、弱い一貫性管理方法を用いて一貫性管理を行い、かつ、主サイトから各サイトへの更新伝播を遅延させて行う (S 3 0 6)。



【特許請求の範囲】

【請求項1】 任意のオブジェクトに対する複製データをそれぞれ有する第1のサイトと、前記複数の第1のサイトにおける複製データの一貫性を集中管理する第2のサイトと、を備えた分散型データベースシステムの一貫性管理方法において、
予め前記複製データに対するリードまたは更新を含む各トランザクションの種類に、一貫性管理のレベルを表す第1レベル、第2レベルおよび第3レベルのいずれかを設定する一貫性レベル設定工程と、
システム内においてトランザクションが発生した場合に、発生したトランザクションの種類に基づいて、前記一貫性管理のレベルを判定する判定工程と、
前記判定工程で第1レベルであると判定された場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、第2のサイトから複数の第1のサイトへの更新伝播を即時行う第1の一貫性管理工程と、
前記判定工程で第2レベルであると判定された場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、第2のサイトから複数の第1のサイトへの更新伝播を遅延させて行う第2の一貫性管理工程と、
前記判定工程で第3レベルであると判定された場合に、弱い一貫性管理方法を用いて一貫性管理を行い、かつ、第2のサイトから複数の第1のサイトへの更新伝播を遅延させて行う第3の一貫性管理工程と、
を含むことを特徴とする分散型データベースシステムの一貫性管理方法。

【請求項2】 任意のオブジェクトに対する複製データをそれぞれ有する第1のサイトと、前記複数の第1のサイトにおける複製データの一貫性を集中管理する第2のサイトと、を備えた分散型データベースシステムの一貫性管理方法において、
予め前記複製データに対するリードまたは更新を含む各トランザクションの種類に、一貫性管理のレベルを表す第1レベル、第2レベルおよび第3レベルのいずれかを設定する一貫性レベル設定工程と、
システム内においてトランザクションが発生した場合に、発生したトランザクションの種類に基づいて、前記一貫性管理のレベルを判定する判定工程と、
前記判定工程で第1レベルであると判定された場合に、
read-onlyのトランザクションまたは更新を含むトランザクションのいずれであっても、厳格な一貫性管理方法を用いて一貫性管理を行い、第2のサイトから複数の第1のサイトへの更新伝播を即時行う第1の一貫性管理工程と、
前記判定工程で第2レベルであると判定された場合に、
read-onlyのトランザクションは緩和された一貫性管理方法を用いて一貫性管理を行い、更新を含むトランザクションは厳格な一貫性管理方法を用いて一貫性管理を行い、第2のサイトから複数の第1のサイトへの

更新伝播を遅延させて行う第2の一貫性管理工程と、
前記判定工程で第3レベルであると判定された場合に、
read-onlyのトランザクションまたは更新を含むトランザクションのいずれであっても、弱い一貫性管理方法を用いて一貫性管理を行い、第2のサイトから複数の第1のサイトへの更新伝播を遅延させて行う第3の一貫性管理工程と、
を含むことを特徴とする分散型データベースシステムの一貫性管理方法。

10 【請求項3】 前記第2および第3の一貫性管理工程における更新伝播の遅延とは、第2のサイトにおいて予め設定されている更新伝播条件が満足された場合に更新伝播を行うことであることを特徴とする請求項1または2に記載の分散型データベースシステムの一貫性管理方法。

【請求項4】 前記第2および/または第3の一貫性管理工程における更新伝播の遅延とは、前記第1のサイトにおいて前記トランザクション毎に更新伝播を行うタイミングを定めた更新伝播条件が設定され、第2のサイトにおいて前記更新伝播条件を満足するように複数の第1のサイトへの更新伝播を行うことであることを特徴とする請求項1～3のいずれか一つに記載の分散型データベースシステムの一貫性管理方法。

【請求項5】 前記第2および/または第3の一貫性管理工程は、少なくとも通信回線および放送波を用いて第2のサイトから複数の第1のサイトへの更新伝播を行うことができ、

前記更新伝播条件は、前記通信回線および放送波のいずれを用いて更新伝播を行うかについての指定を含むことを特徴とする請求項4に記載の分散型データベースシステムの一貫性管理方法。

【請求項6】 前記判定工程は、前記一貫性管理のレベルを判定することに加えて、データの新規登録か否かを判定し、前記新規登録であると判定した場合に、設定されているレベルに関係なく、前記第3レベルと判定することを特徴とする請求項1～5のいずれか一つに記載の分散型データベースシステムの一貫性管理方法。

【請求項7】 前記請求項1～6のいずれか一つに記載の分散型データベースシステムの一貫性管理方法の各工程をコンピュータに実行させるためのプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、分散型データベースシステムにおける複製データの一貫性管理方法に関し、より詳細には、一貫性管理のレベルを複数段階に分け、トランザクションの種類に応じて異なる一貫性管理を行うことができるようにすることにより、分散型データベースシステムの利便性の向上を図った分散型データ

ベースシステムの一貫性管理方法およびその方法の各工程をコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体に関する。

【 0 0 0 2 】

【従来の技術】分散型データベースシステムにおいては、同一データについての複製データが複数のサイトにそれぞれ分散しており、かつ、複数のトランザクションが並行して実行されることになるため、システム上に存在する複製データ間の一貫性をいかに保つかが重要な課題となっている。

【 0 0 0 3 】分散型データベースシステムにおける一貫性管理方法の一例として、データの変更を直列化する操作を行う強い一貫性管理と呼ばれる方法がある。この方法によれば、直列化可能性を保証するが故に操作の並列性を低下させ、一貫性管理のコストが増大するという短所があるものの、直列化可能性が保証されるため、複製データ間の矛盾が一切生じないという長所がある。この強い一貫性管理は、銀行のOLTP (On-Line Transaction Processing) 等に適用されていることが多い。

【 0 0 0 4 】また、一貫性管理方法の他の例として、複製データ間に矛盾が生じることを認める弱い一貫性管理と呼ばれる方法がある。この弱い一貫性管理では、一つのトランザクションの中でシステム全体の一貫性を保証するのではなく、別のトランザクションで更新伝播を行う。この方法によれば、複製データ間に矛盾が生じてしまうことがあるという短所があるものの、適用する応用によってはその矛盾も問題とならない場合があり、一貫性保持に要するコストを軽減することができるという長所がある。この弱い一貫性管理は、データの更新が少な

い安定したシステムに適用されていることが多い。

【 0 0 0 5 】なお、弱い一貫性管理方式については各種の研究がなされており、大別すると、(1) マルチバージョン方式 (Multi-version Control)、(2) 差異限定管理方式 (Bounded Divergence Control)、および (3) 意味ベースの方式 (Semantic-based Control) に分類することができる。したがって、分散型データベースシステムの規模やデータ内容を考慮して、どのような弱い一貫性管理方式を利用すべきかを決定する必要がある。

【 0 0 0 6 】

【発明が解決しようとする課題】しかしながら、従来の分散型データベースシステムの一貫性管理方法においては、上述した強い一貫性管理および弱い一貫性管理のいずれか一方しか利用することができないため、複製データの一貫性管理において不便な場合があるという問題点があった。この問題点は、分散型データベースシステムの規模・データ内容に応じて、トランザクションの種類毎に強い一貫性管理および弱い一貫性管理を使い分けす

ることができる方が便利であるという要求に基づくものである。

【 0 0 0 7 】また、弱い一貫性管理を採用した従来の分散型データベースシステムの一貫性管理方法においては、一つのトランザクションの中でシステム全体の一貫性を保証するのではなく、別のトランザクションで更新伝播を行うため、更新伝播がいつ行われるかユーザは認識することができないという問題点があった。すなわち、更新伝播を行うタイミングはシステムに依存しており、ユーザの所望するタイミングで更新伝播を行うことができるようにすることはできなかった。

【 0 0 0 8 】本発明は上記に鑑みてなされたものであって、分散型データベースシステムの一貫性管理のレベルを複数段階に分け、トランザクションの種類に応じて異なる一貫性管理を行うことができるようにすることにより、分散型データベースシステムの利便性の向上を図ることを第 1 の目的とする。

【 0 0 0 9 】本発明は上記に鑑みてなされたものであって、弱い一貫性管理を採用した場合に、更新伝播を行うタイミングを設定できるようにすることにより、分散型データベースシステムの利便性のさらなる向上を図ることを第 2 の目的とする。

【 0 0 1 0 】

【課題を解決するための手段】上記目的を達成するため、請求項 1 の分散型データベースシステムの一貫性管理方法は、任意のオブジェクトに対する複製データをそれぞれ有する第 1 のサイトと、前記複数の第 1 のサイトにおける複製データの一貫性を集中管理する第 2 のサイトと、を備えた分散型データベースシステムの一貫性管理方法において、予め前記複製データに対するリードまたは更新を含む各トランザクションの種類に、一貫性管理のレベルを表す第 1 レベル、第 2 レベルおよび第 3 レベルのいずれかを設定する一貫性レベル設定工程と、システム内においてトランザクションが発生した場合に、発生したトランザクションの種類に基づいて、前記一貫性管理のレベルを判定する判定工程と、前記判定工程で第 1 レベルであると判定された場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、第 2 のサイトから複数の第 1 のサイトへの更新伝播を即時行う第 1 の一貫性管理工程と、前記判定工程で第 2 レベルであると判定された場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、第 2 のサイトから複数の第 1 のサイトへの更新伝播を遅延させて行う第 2 の一貫性管理工程と、前記判定工程で第 3 レベルであると判定された場合に、弱い一貫性管理方法を用いて一貫性管理を行い、かつ、第 2 のサイトから複数の第 1 のサイトへの更新伝播を遅延させて行う第 3 の一貫性管理工程と、を含むものである。

【 0 0 1 1 】また、請求項 2 の分散型データベースシステムの一貫性管理方法は、任意のオブジェクトに対する

複製データをそれぞれ有する第 1 のサイトと、前記複数の第 1 のサイトにおける複製データの一貫性を集中管理する第 2 のサイトと、を備えた分散型データベースシステムの一貫性管理方法において、予め前記複製データに対するリードまたは更新を含む各トランザクションの種類に、一貫性管理のレベルを表す第 1 レベル、第 2 レベルおよび第 3 レベルのいずれかを設定する一貫性レベル設定工程と、システム内においてトランザクションが発生した場合に、発生したトランザクションの種類に基づいて、前記一貫性管理のレベルを判定する判定工程と、前記判定工程で第 1 レベルであると判定された場合に、`read-only`のトランザクションまたは更新を含むトランザクションのいずれであっても、厳格な一貫性管理方法を用いて一貫性管理を行い、第 2 のサイトから複数の第 1 のサイトへの更新伝播を即時行う第 1 の一貫性管理工程と、前記判定工程で第 2 レベルであると判定された場合に、`read-only`のトランザクションは緩和された一貫性管理方法を用いて一貫性管理を行い、更新を含むトランザクションは厳格な一貫性管理方法を用いて一貫性管理を行い、第 2 のサイトから複数の第 1 のサイトへの更新伝播を遅延させて行う第 2 の一貫性管理工程と、前記判定工程で第 3 レベルであると判定された場合に、`read-only`のトランザクションまたは更新を含むトランザクションのいずれであっても、弱い一貫性管理方法を用いて一貫性管理を行い、第 2 のサイトから複数の第 1 のサイトへの更新伝播を遅延させて行う第 3 の一貫性管理工程と、を含むものである。

【0012】また、請求項 3 の分散型データベースシステムの一貫性管理方法は、請求項 1 または 2 に記載の分散型データベースシステムの一貫性管理方法において、前記第 2 および第 3 の一貫性管理工程における更新伝播の遅延について、第 2 のサイトにおいて予め設定されている更新伝播条件が満足された場合に更新伝播を行うこととしたものである。

【0013】また、請求項 4 の分散型データベースシステムの一貫性管理方法は、請求項 1 ～ 3 のいずれか一つに記載の分散型データベースシステムの一貫性管理方法において、前記第 2 および／または第 3 の一貫性管理工程における更新伝播の遅延とは、前記第 1 のサイトにおいて前記トランザクション毎に更新伝播を行うタイミングを定めた更新伝播条件が設定され、第 2 のサイトにおいて前記更新伝播条件を満足するように複数の第 1 のサイトへの更新伝播を行うこととしたものである。

【0014】また、請求項 5 の分散型データベースシステムの一貫性管理方法は、請求項 4 に記載の分散型データベースシステムの一貫性管理方法において、前記第 2 および／または第 3 の一貫性管理工程が、少なくとも通信回線および放送波を用いて第 2 のサイトから複数の第 1 のサイトへの更新伝播を行うことができ、前記更新伝

播条件が、前記通信回線および放送波のいずれを用いて更新伝播を行うかについての指定を含むものである。

【0015】また、請求項 6 の分散型データベースシステムの一貫性管理方法は、請求項 1 ～ 5 のいずれか一つに記載の分散型データベースシステムの一貫性管理方法において、前記判定工程が、前記一貫性管理のレベルを判定することに加えて、データの新規登録か否かを判定し、前記新規登録であると判定した場合に、設定されているレベルに関係なく、前記第 3 レベルと判定するものである。

【0016】さらに、請求項 7 のコンピュータ読み取り可能な記録媒体にあっては、前記請求項 1 ～ 6 のいずれか一つに記載の分散型データベースシステムの一貫性管理方法の各工程をコンピュータに実行させるためのプログラムを記録したものである。

【0017】

【発明の実施の形態】以下、本発明に係る分散型データベースシステムの一貫性管理方法およびその方法の各工程をコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体の実施の形態について、添付の図面を参照しつつ詳細に説明する。

【0018】〔実施の形態 1〕実施の形態 1 に係る分散型データベースシステムの一貫性管理方法においては、複製データを有するサイトが超多地点に存在する分散型データベースシステムに適用することを想定している。

【0019】ここで、実施の形態 1 に係る分散型データベースシステムの一貫性管理方法を説明するに先立って、分散型データベースシステムを構築するために考慮すべき二つの条件について説明する。

【0020】一つ目は、システム内に存在する複製データの更新を行うことができるサイトはどこかということであり、これには集中方式と分散方式とがある。集中方式は、分散型データベースシステムにおける複数のサイトを主サイトと主サイト以外のサイトに分類し、主サイトでのみシステム内に存在する複製データの更新を可能とし、他のサイトではその更新を行うことができないようにするものである。一方、分散方式は、どのサイトでもシステム内の複製データの更新を可能とし（Any site Writable）、更新を行うサイトが全ての複製データとの調整を行うというものである。

【0021】二つ目は、一時的なデータ矛盾を認めるか否かということであり、これには従来技術で説明した強い一貫性管理と弱い一貫性管理とがある。強い一貫性管理は、システム内に存在する複製データ間の矛盾を認めず、一つのトランザクションの中で全ての複製データ間の整合性を保証するものである。例えば、2 相コミットプロトコルがこの方式に該当する。一方、弱い一貫性管理は、応用の内容に特化して考えることにより、システム内に存在する複製データ間の矛盾を一時的に認めるものである。この方式では、一つのトランザクションの中

で全ての複製データ間の一貫性が保証されるのではなく、更新の伝播は別のトランザクションで行われる。

【0022】図1は、集中管理または分散管理と、強い一貫性管理または弱い一貫性管理との組み合わせによって、分散型データベースシステムを構築した場合の長所および短所が変化することを示した説明図である。実施の形態1に係る分散型データベースシステムの一貫性管理方法においては、後述するようにトランザクションの種類に応じて強い一貫性管理および弱い一貫性管理を設定できるようにするために、プロトコルが簡単な集中管理方式を採用することにする。

【0023】図2は、実施の形態1に係る分散型データベースシステムの一貫性管理方法を実現するための分散型データベースシステムの概念構成図である。図2に示す分散型データベースシステムは、複数のサイトからなり、複数のサイトは、任意のオブジェクトから生成した複製データをそれぞれ有するサイトDBS₁、サイトDBS₂、サイトDBS₃、・・・、サイトDBS_nと、複製データの一貫性を集中管理する主サイトDBS₀と、から構成されている。なお、以下の説明において、サイトDBS₁、サイトDBS₂、サイトDBS₃、・・・、サイトDBS_nをまとめて示す場合にはサイトDBSと記述するものとする。

【0024】ここで、主サイトDBS₀が管理する情報には、システム内に存在する複製データに対する各オブジェクト毎に、対象オブジェクトID（いずれのオブジェクトに対する複製データかを特定するためのものであり、各サイトDBSが保有する複製データにも共通のものが付与されている）、オブジェクトの最新値、最終更新タイムスタンプ、最終更新サイトID、最終更新トランザクションID等がある。一方、各サイトDBSは、システム内に存在する全てのオブジェクトの複製データを有していなければならないというわけではない。さらに、各オブジェクトのマスターデータについては、主サイトDBS₀が集中管理しても良いし（主サイトDBS₀が管理している「オブジェクトの最新値」をマスターデータととらえても良い）、各サイトDBS毎に管理することにしても良い。

【0025】なお、図2には、主サイトDBS₀の下にサイトDBSを設けた様子のみが示されているが、各サイトDBSの下にもさらに複数のサイトDBSを設けて階層構造を形成することもできる。このような場合においては、上位のサイトDBSを下位のサイトDBSの主サイトとして動作させることもできる。

【0026】また、実施の形態1に係る分散型データベースシステムの一貫性管理方法においては、一貫性管理のレベルとしてレベルA～Cを用意し、トランザクションの種類に応じて、異なるレベルの一貫性管理を行うことができるようにしている。

【0027】レベルA（本発明の第1レベルに該当す

る）は、read-onlyのトランザクションまたは更新を含むトランザクションのいずれであっても、厳格な一貫性管理方法を用いて複製データの一貫性管理を行い、主サイトDBS₀から各サイトDBSへの更新伝播を即時行うというものである。

【0028】換言すれば、レベルAは、一つのトランザクションの中で各サイトDBSへの更新伝播まで実行することによって複製データの一貫性管理を行うものである。トランザクションの直列可能性を保証するものである。

【0029】また、レベルB（本発明の第2レベルに該当する）は、read-onlyのトランザクションについては緩和された一貫性管理方法を用いて複製データの一貫性管理を行い、更新を含むトランザクションについては厳格な一貫性管理方法を用いて複製データの一貫性管理を行い、主サイトDBS₀から各サイトDBSへの更新伝播を遅延させて行うというものである。

【0030】レベルBにおいては、ある複製データの更新処理を行う場合に、複数のトランザクションに分散して、主サイトDBS₀から各サイトDBSへの更新伝播が行われる。データアクセスに際しては、1コピー直列化可能性は保証する（P. A. Bernstein, V. Hadzilacos, and N. Goodman: Concurrency Control and Recovery in Database Systems, Addison-Wesley, 1987）。1コピー直列化可能である場合、データ更新の際には同一データに対する複数の複製データの中から最新バージョンを探し、同一データに対する更新処理が競合しないように処理するが、読み出しだけの際には、多少データが古くても良いのでローカルデータをそのまま読むという処理を行う。

【0031】さらに、レベルC（本発明の第3レベルに該当する）は、read-onlyのトランザクションまたは更新を含むトランザクションのいずれであっても、弱い一貫性管理方法を用いて一貫性管理を行い、主サイトDBS₀から各サイトDBSへの更新伝播を遅延させて行うというものである。

【0032】レベルCは、複製データ間の一時的な一貫性の崩壊を認め、各サイトDBS毎に複製データの更新を可能とするものである。更新結果は複数のトランザクションとなって主サイトDBS₀から各サイトDBSへ伝播される。この場合、同一データに対する複製データについての更新の競合（衝突）が検出された場合の解決法が問題となる。レベルCにおいては、後述するように、コミットした後に更新の競合を検出する処理が行われるため、通常の方法のアボートはできない。更新の競合が生じた場合については、後に説明することにする。

【0033】なお、上述したレベルA～Cは、後述するように、read-onlyまたは更新を含むトランザ

10

20

30

40

50

クシヨンの種類に応じて設定される。

【0034】続いて、実施の形態1に係る分散型データベースシステムの一貫性管理方法を具体的に説明する。図3(a)は一貫性管理のレベルを設定する手順を、図3(b)は図3(a)に示すようにして設定した一貫性管理のレベルに基づいて一貫性管理方法を実行する手順を示すフローチャートである。ここでは、図3(a)および図3(b)を用いて一貫性管理手順の概略を説明した後、レベルA～Cの一貫性管理方法について具体的に説明する。

【0035】(一貫性管理手順の概略) まず、図3

(a)に示すように、分散型データベースシステムの規模・データ内容等を考慮し、システム内で発生するトランザクションの種類に応じて予め一貫性管理のレベルを設定する(S300:本発明の一貫性レベル設定工程に該当する)。すなわち、各トランザクションの種類に対して、レベルA～Cが設定される。なお、一貫性管理のレベルは、分散型データベースシステムのセットアップ時に設定することができ、また、トランザクションの発生時に設定することもできる。

【0036】図3(a)のステップS300においてレベルA～Cを設定した後、図3(b)に示すように、分散型データベースシステム内にトランザクションが発生すると(S301)、発生したトランザクションに設定されている一貫性管理のレベルの判定処理を実行する(S302:本発明の判定工程に該当する)。

【0037】そして、ステップS302における判定処理の結果に基づいて、発生したトランザクションに設定された一貫性管理のレベルがレベルA、レベルBおよびレベルCのいずれであるかを判定する(S303:本発明の判定工程に該当する)。ステップS303において、レベルAであると判定した場合には、レベルAの一貫性管理方法を実行し(S304:本発明の第1の一貫性管理工程に該当する)、レベルBであると判定した場合には、レベルBの一貫性管理方法を実行し(S305:本発明の第2の一貫性管理工程に該当する)、さらに、レベルCであると判定した場合には、レベルCの一貫性管理方法を実行する(S306:本発明の第3の一貫性管理工程に該当する)。

【0038】(レベルAの一貫性管理方法) 次に、レベルAによる一貫性管理方法について、強い一貫性管理方法の概略について説明した後、実施の形態1のレベルAによる強い一貫性管理方法を具体的に説明することにする。

【0039】強い一貫性管理方法としては、2相ロック、楽観的2相ロック、グローバルタイムスタンプに基づく楽観的方法、ベーシックタイムスタンプオーダリング等の各種があり、これらによって一つのトランザクションの中で全ての複製データを更新する場合には直列化可能となる。なお、「完全に複製されたシステムでは、

楽観的2相ロックが最も適している」と述べている論文がある(M. J. Carey and M. Livny: "Conflict detection and recovery for replicated data," ACM TODS, Vol. 16, No. 4, pp. 703-746, Dec. 1991:シミュレーションでパフォーマンス評価を行った代表的論文である)。

【0040】上記強い一貫性管理方法のうち、楽観的制御方式ではコミット時までロックを行わずにプライベート領域(ワークエリア)にのみ複製データの更新結果を保存しておき、コミット時になって初めてトランザクション間の競合があったかどうかを調べる。

【0041】ここで、あるデータに対する複数の複製データがシステム内に存在する場合の楽観的2相ロック方式の代表例な処理の例を以下に示す(1CDCS95のJingのアルゴリズム)。

(1) readとwriteに対しては、全複製データX0, X1, ..., XnをRlock(readモードのロック)する(Rlockはトランザクションが複製データを利用していることを示すためのものである)。

(2) トランザクションの終了時に、writeした複製データをWlock(writeモードのロック)する。Wlockできればトランザクションをコミットし、そうでなければアボートする。

(3) 全ロックを解除する。

【0042】上記楽観的2相ロック方式においては、確認相で全てのサイトDBSにロックをかけなければならないことが問題である。すなわち、実際には全てのサイトDBSにロックをかけることは難しいため、楽観的2相ロック方式は障害に弱いといえる。

【0043】その理由について、集中管理に基づく2相ロック方式と比較して説明する。集中管理に基づく2相ロック方式の場合は、トランザクションがreadしたとき、マスターデータに対してRlockをかけてから、ローカルデータを読み出す。writeの場合も同様で、マスターデータにWlockをかけてから全複製データをwriteする。したがって、マスターデータのみをロックすれば良いため、負荷を軽減できる。ところが、楽観的2相ロック方式の場合には、全てのサイトでwriteできなければアボートしなくてはならないため、障害に弱いといえる。

【0044】また、ダウンしている(参加できない)サイトがあっても、処理が続けられるようにするために、以下の2つの方式が考えられる。

【0045】第1の方式は、重み付き投票方式(Giffordによって提案されたもの)である。重み付き投票方式においては、以下の条件を満たす一定数のサイトが参加すれば、その中に必ず最新版の値が含まれること

が保証されている。

Read-quorum+Write-quorum>n

Write-quorum+Write-quorum>n

ここで、nは複製データの数を表す。

【0047】第2の方式は、集中管理に基づくタイムスタンプベースの楽観的方式(PTO: Primary-based Timestamp Optimistic Control)である。PTO方式は集中管理方式であるため、ダウンしているサイトについては無視することができる。

【0048】具体的に、集中管理方式では、更新が承認された複製データの最新の値が主サイトDBS_rに保持される。よって楽観的制御方式の確認相(validation phase)において一貫性のチェックを行う場合に主サイトDBS_rに問い合わせるだけで済み、他の複製データにWlockをかけにいく必要はなくなる。切断していたサイトが再接続した場合の制御も同じで良い。PTO方式において複数種類の複製データにアクセスするトランザクションにあっては、ローカルサイトにそれらの複製を作成してから実行を始めることとする。

【0049】実施の形態1に係る分散型データベースシステムの一貫性管理方法を超多地点にわたる分散型データベースシステムに適用することを想定した場合には、重み付き投票方法よりPTO方式の方が優れていると考えられる。なぜなら、PTO方式の場合、一貫性のチェックを行う場合に主サイトDBS_rに問い合わせるだけで済むため、主サイトDBS_rがダウンしない限り、通信コストの削減を図ることができるからである。

【0050】そこで、実施の形態1に係る分散型データベースシステムの一貫性管理方法においては、レベルAの一貫性管理方法としてPTO方式を採用することにし、以下にレベルAの一貫性管理方法を具体的に説明する。ここでは、サイトDBS_iでread-onlyまたはread-only以外の更新を含むトランザクションTが発生するものとする。

【0051】なお、以下の説明において、複製データはオブジェクトOに対してn個あると仮定し、それをO₁, ..., O_nと表すと共に、主サイトDBS_rが有するオブジェクトOの複製データをO₀と表す。オブジェクトOの最後に更新された時刻を表すタイムスタンプをt(O)とし、またトランザクションTが保持するオブジェクトOの最終更新タイムスタンプをt(O, T)と表すことにする。

【0052】サイトDBS_iは、トランザクションTが発生すると、そのトランザクションTに設定されている一貫性管理のレベルを判定する処理を行う。発生したトランザクションTがレベルAの場合、レベルAの一貫性管理方法が実行される。

【0053】(1) 更新を含むトランザクションの場合

【0046】

① 読み出し相

サイトDBS_iは、ローカルデータ、即ち複製データO₀にアクセスし、複製データO₀をプライベート領域(ワークエリア)に読み出す。そして、プライベート領域に読み出された複製データO₀についてreadおよびwriteが行われる。複製データO₀をプライベート領域に読み出す際には、複製データO₀の最後に更新された時刻を表すタイムスタンプt(O₀)をトランザクションTが保持する複製データO₀の最終更新タイムスタンプに設定し、t(O₀, T)とする。

【0054】② 確認相

サイトDBS_iは、主サイトDBS_rに対してレベルAの一貫性管理方法の要求を含むREQUESTメッセージを送信する。図4は、REQUESTメッセージを示す説明図であり、図4に示すデータがREQUESTメッセージに設定され、主サイトDBS_rに対して送信される。なお、REQUESTメッセージには、メッセージの送信元であるサイトDBSのサイトID、トランザクションID、さらに、トランザクションTがアクセスした複製データO_i毎に、操作(readまたはwrite)、対象オブジェクトID、タイムスタンプt(O_i, T)および更新後の値が設定される。

【0055】主サイトDBS_rは、REQUESTメッセージを受信し、トランザクションTにレベルAが設定されていると判定して、受信したREQUESTメッセージを待ち行列に入力する。待ち行列に入力されたREQUESTメッセージは、入力された順番で処理される。

【0056】ここで、主サイトDBS_rは、サイトDBS_iによってアクセスされた複製データO₀の全てについて、

【数1】

$$t(O_i, T) \geq t(O_0)$$

の条件を満たすか否かを危険領域(critical section)内で判定する。

【0057】上記条件を満たすと判定した場合、サイトDBS_iで更新された複製データO₀に該当する複製データO₀をREQUESTメッセージ中の更新後の値を用いて更新し(更新後の値を保持する)、そのタイムスタンプをt(O₀) = Clock(現在時刻のタイムスタンプ)に変更した後、図5に示すコミットメッセージを全てのサイトDBS(サイトDBS_i, サイトDBS_j, ..., サイトDBS_n:更新された複製データO_iに該当する複製データを有するサイトDBS)に送信する。このコミットメッセージは、REQUESTメッセージに基づいて生成され、REQU

ESTメッセージの送信元であるサイトDBSのサイトID、トランザクションID、さらに、複製データ O_i 毎に、対象オブジェクトID、タイムスタンプ $t(O_i)$ および更新後の値が設定される。

【0058】すなわち、更新を含むトランザクションの場合、主サイトDBS_rはそのトランザクション内でサイトDBS₁、サイトDBS₂、サイトDBS₃、・・・、サイトDBS_nの全てに更新情報の伝播を行う。ただし、2相コミット方式とは異なり、サイトDBS₁、サイトDBS₂、サイトDBS₃、・・・、サイトDBS_nからの返事は必要としない。つまり、ダウンしているサイトDBSがあっても無視して処理を進めることができる。

【0059】なお、ダウンしているサイトDBSが復旧した場合、そのサイトDBSは、主サイトDBS_rに対して何らかのトランザクションを実行することにより、最新の更新状態を知ることができる。また、復旧した後に、主サイトDBS_rに対して最新の更新状態を問い合わせることができるようにすることもできる。

【0060】一方、上記条件を満たさないと判定した場合、図6に示すアボートメッセージをサイトDBS_iに送信する。このアボートメッセージは、REQUESTメッセージに基づいて生成され、REQUESTメッセージの送信元であるサイトDBSのサイトID、トランザクションID、さらに、複製データ O_i 毎に、対象オブジェクトID、タイムスタンプ $t(O_i)$ および現在の値 O_i が設定される。

【0061】㊸ 書き込み相
サイトDBS₁、サイトDBS₂、サイトDBS₃、・・・、サイトDBS_nは、コミットメッセージを受信した場合、コミットメッセージに含まれている更新後の値およびそのタイムスタンプ値を用いて該当する複製データを更新する。

【0062】一方、サイトDBS_iは、アボートメッセージを受信した場合、トランザクションTをアボートし、アボートメッセージに含まれている最新値およびタイムスタンプ値を用いて複製データ O_i およびそのタイムスタンプ $t(O_i)$ を更新する。

【0063】(2) read-onlyのトランザクションの場合

続いて、read-onlyのトランザクションの場合について説明する。サイトDBS₁は、ローカルデータ、即ち複製データ O_i にアクセスし、複製データ O_i をプライベート領域（ワークエリア）に読み出す。ここで、複製データ O_i をプライベート領域に読み出す際には、複製データ O_i のタイムスタンプ $t(O_i)$ をトランザクションTが保持する複製データ O_i のタイムスタンプに設定し、 $t(O_i, T)$ とする。

【0064】サイトDBS₁は、主サイトDBS_rに対してレベルAの一貫性管理方法の要求を含むREQUE 50

STメッセージを送信する。このREQUESTメッセージは図4に示したものと同様であるが、read-onlyのトランザクションであるため、更新後の値は除かれる。

【0065】主サイトDBS_rは、REQUESTメッセージを受信し、トランザクションTにレベルAが設定されていると判定して、受信したREQUESTメッセージを待ち行列に入力する。待ち行列に入力されたREQUESTメッセージは順番に処理される。

【0066】主サイトDBS_rは、サイトDBS_iによってアクセスされた複製データ O_i について、
【数2】

$$t(O_i, T) \geq t(O_i)$$

の条件を満たすか否かを危険領域（critical section）内で判定する。

【0067】上記条件を満たすと判定した場合、図5に示したコミットメッセージをサイトDBS_iに送信する。ただし、read-onlyのトランザクションであるため、更新後の値は除かれる。一方、上記条件を満たさないと判定した場合、図6に示したアボートメッセージをサイトDBS_iに送信する。

【0068】サイトDBS_iは、コミットメッセージを受信した場合、厳格に一貫性が保たれているとしてread-onlyのトランザクションをコミットする。一方、サイトDBS_iは、アボートメッセージを受信した場合、read-onlyのトランザクションをアボートし、アボートメッセージに含まれている最新値およびタイムスタンプ値を用いて複製データ O_i およびそのタイムスタンプ $t(O_i)$ を更新する。

【0069】（レベルBの一貫性管理方法）つぎに、レベルBによる一貫性管理方法について説明する。レベルAでは一つのトランザクションの中で全てのサイト（サイトDBS₁、サイトDBS₂、サイトDBS₃、・・・、サイトDBS_n）への更新を行うことにしているが、レベルBでは、別のトランザクションで更新を伝播させることにする。つまり、即時に更新を伝播させるのではなく、後述する各種の差異限定制約条件（本発明の更新伝播条件に該当する）を設け、その条件が満足された場合に更新伝播を行うという処理を行う。

【0070】ここでは、サイトDBS₁でread-onlyまたはread-only以外の更新を含むトランザクションTが発生するものとして、レベルBの一貫性管理方法を具体的に説明する。

【0071】サイトDBS₁は、トランザクションTが発生すると、そのトランザクションTに設定されている一貫性管理のレベルを判定する処理を行う。発生したトランザクションTがレベルBの場合、レベルBの一貫性管理方法が実行される。

【0072】(1) 更新を含むトランザクションの場合

① 読み出し相

サイトDBS_iは、ローカルデータ、即ち複製データO_iにアクセスし、複製データO_iをプライベート領域(ワークエリア)に読み出す。そして、プライベート領域に読み出された複製データO_iについてreadおよびwriteが行われる。ここで、複製データO_iをプライベート領域に読み出す際には、複製データO_iの最後に更新された時刻を表すタイムスタンプt(O_i)をトランザクションTが保持する複製データO_iの最終更新タイムスタンプに設定し、t(O_i, T)とする。

【0073】② 確認相

サイトDBS_iは、主サイトDBS_rに対してレベルBの一貫性管理方法の要求を含むREQUESTメッセージを送信する。このREQUESTメッセージは、図4に示したものと同様のものである。

【0074】主サイトDBS_rは、REQUESTメッセージを受信し、トランザクションTにレベルBが設定されていると判定して、受信したREQUESTメッセージを待ち行列に入力する。待ち行列に入力されたREQUESTメッセージは、入力された順番で処理される。

【0075】ここで、主サイトDBS_rは、サイトDBS_iによってアクセスされた複製データO_iの全てについて、

【数3】

$$t(O_i, T) \geq t(O_0)$$

の条件を満たすか否か危険領域(critical section)内で判定する。

【0076】上記条件を満たすと判定した場合、サイトDBS_iで更新された複製データO_iに該当する複製データO₀をREQUESTメッセージ中の更新後の値に更新し(更新後の値を保持する)、そのタイムスタンプをt(O₀)=Clock(現在時刻のタイムスタンプ)に変更した後、図5に示したコミットメッセージをサイトDBS_iに送信する。加えて、主サイトDBS_rは、上記トランザクションで更新された更新情報をキューに入れておき、差異限定制約条件が満足された場合に、この更新情報を別のトランザクションにおいて他のサイトDBS(サイトDBS₁, サイトDBS₂, ..., サイトDBS_n)に送信することによって更新伝播を行う。なお、REQUESTメッセージを送信して来たサイトDBS(ここではサイトDBS_i)に対しては、通信コストを下げるために更新伝播を行わないようにする。

【0077】一方、上記条件を満たさないと判定した場合、図6に示したアポートメッセージをサイトDBS_iに送信する。

【0078】ここで、上記差異限定制約条件について説明する。差異限定制約条件としては、例えば、以下の

a) ~ e) 等を挙げることができる。

【0079】a) リフレッシュする周期を設定し、設定した周期毎に更新伝播を行う。主サイトDBS_rは、一定の周期毎に、他の全てのサイトDBS(サイトDBS₁, サイトDBS₂, ..., サイトDBS_n)に更新情報を送信することにより、更新伝播を行う。主サイトDBS_rは、データ更新情報をキューに入れておき、周期毎にそれを送り出す。受け側のサイトDBS(サイトDBS₁, サイトDBS₂, ..., サイトDBS_n)はこの更新情報が送られてくることにより、主サイトDBS_rがダウンしていないことを知ることができる。

【0080】b) キューにたまっている更新情報の数が一定数を超えた場合に更新伝播を行う。例えば限度数を3とした場合は、更新情報がキューに3つたまったところで更新伝播が行われる。

【0081】c) 更新伝播を行う基準となるデータの領域を決めておき、その領域のデータが更新された場合に更新伝播を行う。例えば、書誌情報のうち、タイトルおよび作者の2つの属性が更新された場合は即時に更新伝播を行うというものであり、直ちに更新情報を公知にする必要があるときに有効である。スキーマの属性として、オブジェクトやその属性に即時更新伝播フラグをつけておき、それがONになっている情報に更新があった場合は即時に更新伝播を行うようにする。ただし、この条件を設定する場合には、他の条件と共に設定する必要がある。

【0082】d) 更新伝播を行う基準となる閾値を設定し、設定した閾値を超えた場合に更新伝播を行う。データの更新の度合いに対して閾値を設定しておく。例えば、コンテンツの1/20以上が更新された場合に更新伝播するというような絶対的な値を設定し、または、参照された回数が100を超えたら即時更新伝播するというような設定を行う。後者の場合、本や番組で人気が出たものはそれを参照する人が多いことを示しているため、即時に情報行進を行うほうが良いという考えに基づくものである。

【0083】e) バージョンが大きく変わった場合に更新伝播を行う。本の改訂版等でマイナーな誤字の直し程度であれば更新伝播を待つようにするが、大幅な修正が加えられた改訂版が出たような場合には、直ちに更新伝播を行うことにするというものである。

【0084】なお、上記a) ~ e)の差異限定制約条件は組み合わせて設定することができる。

【0085】次に、差異限定制約条件を用いた更新伝播手順について説明する。図7(a)は差異限定制約条件を設定する手順を示すフローチャートである。図7

(a)に示すように、システムのセットアップ時において、上記a) ~ e)に示した差異限定制約条件を単独または組み合わせて主サイトDBS_rに設定する(S700: 本発明の更新伝播設定工程に該当する)。

【0086】なお、差異限定制約条件を設定する際には、主サイトDBS₁を操作して設定しても良いし、サイトDBS₁、サイトDBS₂、サイトDBS₃、・・・、サイトDBS_nから設定することもできる。

【0087】また、差異限定制約条件は、上記ステップS700の処理を再び行うことによって後に変更することも可能であるが、図3(a)に示したステップS300において一貫性管理のレベルを設定する際にいずれかの差異限定制約条件を設定しておく必要がある。

【0088】さらに、サイトDBS毎に異なる差異限定制約条件を設定することもできる。例えば、サイトDBS₁については差異限定制約条件a)で更新伝播を行い、サイトDBS₂については差異限定制約条件b)で更新伝播を行うように主サイトDBS₁に対して設定することもできる。この場合、主サイトDBS₁は、各サイトDBS毎に設定された差異限定制約条件を管理し、各サイト毎に異なる差異限定制約条件を用いて更新伝播を行うことになる。

【0089】図7(b)は図7(a)に示すようにして設定した差異限定制約条件に基づいて更新伝播を行う手順を示すフローチャートである。主サイトDBS₁は、常にまたは定期的にキューに入れられた更新情報について、差異限定制約が満足されたか否かを判定する(S701)。

【0090】そして、ステップS701において差異限定制約が満足されたと判定した場合には、キューに入れられた更新情報をサイトDBS₁、サイトDBS₂、サイトDBS₃、・・・、サイトDBS_nに送信することによって更新伝播処理を行う(S702)。

【0091】㊸ 書き込み相
サイトDBS₁は、コミットメッセージを受信した場合、コミットメッセージに含まれている更新後の値およびそのタイムスタンプ値を用いて該当する複製データを更新する。

【0092】また、サイトDBS₁、サイトDBS₂、・・・、サイトDBS_nは、主サイトDBS₁から更新伝播を受信した場合、更新後の値およびそのタイムスタンプ値を用いて該当する複製データを更新する。この際、2相コミット方式とは異なり、サイトDBS₁、サイトDBS₂、サイトDBS₃、・・・、サイトDBS_nからの返事は必要としない。つまり、ダウンしているサイトがあっても無視して処理を進めることができる。

【0093】なお、ダウンしているサイトDBSが復旧した場合、そのサイトDBSは、主サイトDBS₁に対して何らかのトランザクションを実行することにより、最新の更新状態を知ることができる。また、復旧した後に、主サイトDBS₁に対して最新の更新状態を問い合わせることができるようにすることもできる。

【0094】一方、サイトDBS₁は、アボートメッセージを受信した場合、トランザクションをアボートし、

アボートメッセージに含まれている最新値およびタイムスタンプ値を用いて複製データO_iおよびそのタイムスタンプt(O_i)を更新する。

【0095】(2) read-onlyのトランザクションの場合

レベルBの場合は、差異限定制約条件に基づいて更新伝播を行うため、主サイトDBS₁を除き、サイトDBS₁、サイトDBS₂、サイトDBS₃、・・・、サイトDBS_nの複製データは最新バージョンではない可能性が常にある。それをread-onlyのトランザクションにおいて許可するかしないかでレベルAとレベルBとの差が生じる。

【0096】レベルBにおけるread-onlyのトランザクション場合では、ローカルサイト(サイトDBS₁、サイトDBS₂、サイトDBS₃、・・・、サイトDBS_n)の複製データをそのままreadするようにして、直列可能性チェックを緩和する。つまりread-onlyのトランザクションであれば、他のサイトDBSとの調整を行うことなくローカルに実行してしまうことにする。このようにした場合であっても、1コピー直列化は保証される。つまり、他の更新が起こる前にreadしていたと考えることで直列化できる。

【0097】(レベルCの一貫性管理方法) レベルBによる一貫性管理方法は、更新を含むトランザクションに対しては厳格、つまり強い一貫性管理を採用している。これに対し、レベルCによる一貫性管理方法においては、システム全体の処理効率を更に向上させるため、更新を含むトランザクションも緩和して弱い一貫性管理を行うことにする。

【0098】そこで、レベルCの一貫性管理方法を説明する。ここでは、サイトDBS₁でread-onlyまたはread-only以外の更新を含むトランザクションTが発生するものとする。

【0099】サイトDBS₁は、トランザクションTが発生すると、そのトランザクションTに設定されている一貫性管理のレベルを判定する処理を行う。発生したトランザクションTがレベルCの場合、レベルCの一貫性管理方法が実行される。

【0100】(1) 更新を含むトランザクションの場合
㊸ 読み出し相

サイトDBS₁は、ローカルデータ、即ち複製データO_iにアクセスし、複製データO_iをプライベート領域(ワークエリア)に読み出す。そして、プライベート領域に読み出された複製データO_iについてreadおよびwriteが行われる。ここで、レベルCの一貫性管理方法においては、弱い一貫性管理を採用することにより、主サイトDBS₁を介することなく、サイトDBS₁、サイトDBS₂、サイトDBS₃、・・・、サイトDBS_n毎に単独で複製データの更新を行うことができる。そのため、サイトDBS₁は、読み出した複製デー

タO_iを更新し、更新を含むトランザクションをコミットする。

【0101】② 確認相

サイトDBS_iは、上記コミット時に主サイトDBS_r宛てに承認リクエストメッセージ（更新ログ）を送信する。承認リクエストメッセージは、図4に示したREQUESTメッセージと同様のものである。

【0102】主サイトDBS_rは、サイトDBS_iからの承認リクエストメッセージに対して承認または非承認を与える。承認・非承認の判定条件はPTO方式の中で説明した通常のタイムスタンプ方式によるものである。

【0103】ここで、承認・非承認を判定する際に、サイトDBS_iの承認リクエストメッセージと競合する他のサイトDBSの承認リクエストメッセージが存在しない場合には、サイトDBS_iにおいて行われた更新が承認される。その結果、主サイトDBS_rの該当する複製データが更新される（更新後の値を保持し、タイムスタンプを変更）と共に、レベルBで説明したように、その更新情報がキューに入れられ、差異限定制約条件が満足された場合に、この更新情報をサイトDBS（サイトDBS_i、サイトDBS_j、サイトDBS_k、・・・、サイトDBS_n）に送信することによって更新伝播が行われる。なお、差異限定制約条件については、レベルBで説明した通りであるため、ここでは説明を省略する。また、レベルBの場合と同様に、承認リクエストメッセージを送信して来たサイトDBS（ここではサイトDBS_i）に対しては、通信コストを下げるために更新伝播を行わないようにする。

【0104】一方、例えば、サイトDBS_iおよびサイトDBS_jの承認リクエストメッセージが競合したものとする。具体的には、サイトDBS_iおよびサイトDBS_jで同一の複製データO_iを保持していて、両方のトランザクションで同時刻に複製データO_iを更新し、両方のトランザクションがコミットされた後、承認リクエストメッセージ（更新ログ）が主サイトに届いたものとする。主サイトDBS_rは、承認リクエストメッセージが先に届いたサイトの更新を承認し、後から届いた方の更新は承認しない。ここで、サイトDBS_iの承認リクエストメッセージがサイトDBS_jの承認リクエストメッセージより先に主サイトDBS_rに届いたものとする、サイトDBS_iの更新は承認されないことになる。この場合、主サイトDBS_rは、サイトDBS_iに対してアボートメッセージを送信する。

【0105】③ 書き込み相

更新が承認された場合、サイトDBS_iは、既に複製データを更新するためのトランザクションはコミットされているため何ら処理を行う必要はない。一方、他のサイトDBSは、差異限定制約条件によって更新情報が伝播されて来るため、その更新情報に基づいて複製データを更新する。この際、2相コミット方式とは異なり、サイ

トDBS_i、サイトDBS_j、サイトDBS_k、・・・、サイトDBS_nからの返事は必要としない。つまり、ダウンしているサイトがあっても無視して処理を進めることができる。

【0106】なお、ダウンしているサイトDBSが復旧した場合、そのサイトDBSは、主サイトDBS_rに対して何らかのトランザクションを実行することにより、最新の更新状態を知ることができる。また、復旧した後に、主サイトDBS_rに対して最新の更新状態を問い合わせることができるようにすることもできる。

【0107】一方、アボートメッセージを受信した場合は、既にトランザクションはコミットにより完了してしまっているため通常のアボートを行うことはできない。そこで、レベルCにおいては、他のサイトDBSから必要なバージョンの複製データをコピーすることにより更新してしまった複製データを回復することにする。

【0108】ここで、上記承認リクエストメッセージ中の複製データのうち、主サイトDBS_rから更新されている値をコピーする処理、つまりトランザクション自身がアボート用にデータを格納しておき、このデータを用いて更新してしまった複製データを回復するという処理も考えられる。しかしながら、マルチメディアデータは一般にサイズが大きいので、トランザクションの中で複製を作ることはコストがかかり問題となる。また、更新頻度の小さい応用においては更新が衝突する可能性が少ないため、そのような応用においてアボート用にデータの複製を作ることは効率的ではない。

【0109】よって、アボートメッセージを受信した場合の処理として、記憶容量およびコピーに要する時間の点からも他のサイトDBSから必要なバージョンの複製データをコピーして回復する処理を行う方が効率的であるといえる。

【0110】(2) read-onlyのトランザクションの場合

レベルCにおけるread-onlyのトランザクション場合では、ローカルサイト（サイトDBS_i、サイトDBS_j、サイトDBS_k、・・・、サイトDBS_n）の複製データをreadするようにして、直列可能性チェックを緩和する。つまりread-onlyのトランザクションであれば、他のサイトとの調整を行うことなくローカルに実行してしまうことにする。このようにした場合であっても1コピー直列化は保証される。つまり、他の更新が起こる前にreadしていたと考えることで直列化できる。

【0111】ここで、レベルCの応用例として、検索エージェント（人間に代わって複数のDBサイトを縦断検索してくれるようなプログラム）のDBを考える。従来のデータベースはオブジェクト値を永続的に格納することを目的としていた。ところが、検索エージェントのような応用を考えた場合、その記憶としてデータベースを

使うならば、永続的というよりは「少々間違っているけれども良く、間違っていることが判定できれば良い。そして、間違っていたら正しい値に更新すれば良い。」という特徴を持つ。検索サイト名等を格納している場合において、もし、そのサイトが移動したとしても、そのサイトにアクセスしたときに初めて「そのサイトが移動していたこと」がわかれば良い。人間の記憶の変わりにDBを利用とする場合、このような「多少間違ったデータでも良い」という応用例を多く見出すことができる。

【0112】なお、以上説明した一貫性管理のレベルA～Cの特徴を一つにまとめると、図8のようになる。

【0113】このように、実施の形態1に係る分散型データベースシステムの一貫性管理方法によれば、分散型データベースシステムの一貫性管理のレベルとしてレベルA～Cを設け、トランザクションの種類に応じてレベルA～Cによる異なる一貫性管理を行うことにより、複製データの一貫性管理における利便性の向上を図ることができる。

【0114】また、実施の形態1に係る分散型データベースシステムの一貫性管理方法を用いることにより、それぞれ異なる一貫性管理のレベルでトランザクションを実行することができ、トランザクションの種類に応じて最適な複製データの一貫性管理を行うことができる。

【0115】さらに、レベルBおよびレベルCの場合は、差異限定制約が満足された場合に更新伝播を行うことにしたことにより、更新がある毎に更新伝播を行う必要をなくすることができ、複製データの一貫性保持のコストを削減することができる。

【0116】〔実施の形態2〕つぎに、実施の形態2に係る分散型データベースシステムの一貫性管理方法について説明する。例えば、電子図書館等の応用においては、新規の書誌データを一括して登録することが多く、また、大学等でも新入生や新人のデータ登録を一括して行うことが多い。こうした新規登録トランザクションは、既にあるデータを参照しないで独立に処理することが可能であることから、他のトランザクションとの競合の確認処理を省略しても何ら問題は発生せず、むしろ処理の効率化を図る上で好ましい。そのため、実施の形態2に係る分散型データベースシステムの一貫性管理方法においては、実施の形態1で説明した処理に加えて、一貫性管理のレベルを判定する際に新規登録トランザクションか否かを判定し、新規登録トランザクションである場合には、他のトランザクションとの競合の確認処理を省略して、そのまま登録処理を行うことができるようにするものである。

【0117】ところで、楽観的制御においては、読み出し相、確認相および書き込み相からなる処理を行う必要がある。ところが新規登録のみのトランザクションでは、他のトランザクションと競合する可能性は全くないため、トランザクションが競合するか否かの確認を省略

し、読み出し相の処理が終わった後、直接書き込み相の処理を実行することができる。

【0118】そこで、実施の形態2に係る分散型データベースシステムの一貫性管理方法においては、既存のデータと独立なデータを新規登録する場合について、独自のトランザクションレベル（例えば、新規登録処理フラグを設定する）を設け、そのレベルのトランザクションについては一切競合チェックを行わないことにする。

【0119】新規登録トランザクションは、実施の形態1で説明したレベルA～Cのいずれ対しても適用することができる。ただし、効率を考えた場合には、各サイトDBS単位で更新を行うことができるレベルCに適用することが最も優れている。そのため、実施の形態2においては、この新規登録トランザクションをレベルCに適用することを前提として、処理の概略を説明する。図9(a)は一貫性管理のレベルおよび新規登録トランザクションを設定する手順を、図9(b)は図9(a)に示すようにして設定した一貫性管理のレベルおよび新規登録トランザクションの設定に基づいて一貫性管理方法および新規登録を実行する手順を示すフローチャートである。

【0120】まず、図9(a)に示すように、トランザクションの種類に応じて一貫性管理のレベルを設定し（実施の形態1参照）、かつ、データの新規登録のためのトランザクションを新規登録トランザクションとして設定する(S900)。

【0121】図9(a)のステップS900で新規登録トランザクションについての一貫性管理レベルを設定した後、図9(b)に示すように、分散型データベースシステム内にトランザクションが発生すると(S901)、発生したトランザクションに設定されている一貫性管理のレベルの判定処理およびトランザクションが新規登録トランザクションか否かの判定処理を実行する(S902：本発明の判定工程に該当する)。

【0122】そして、ステップS902において、トランザクションが新規登録トランザクションであると判定した場合、設定されているレベルに関係なく、レベルCと判定する(S903：本発明の判定工程に該当する)。そして、ステップS906に進み、レベルCの一貫性管理方法を実行する。

【0123】なお、レベルCで新規登録トランザクションを実行する場合においては、他のサイトDBSにおけるトランザクションとの競合の可能性がないことから、承認リクエストメッセージを用いた競合のチェックは省略される。

【0124】このように、実施の形態2に係る分散型データベースシステムの一貫性管理方法によれば、データの新規登録を行うためのトランザクションを実行する場合において、他のトランザクションとの競合をチェックする処理を省略することにしたため、データを新規登録

する際の処理効率を向上させることができる。

【0125】ただし、INSERT操作の中には既存のデータと意味的に関連付けられているものがある。例えば、readした2つの値の平均値を求め、その値を使ってINSERTするようなトランザクションの場合である。このような場合には、既存のデータと意味的に関連があるため、そのトランザクションは新規登録トランザクションではない。

【0126】〔実施の形態3〕上述したレベルBおよびレベルCの一貫性管理方法においては、主サイトDBS_rに予め設定された差異限定制約条件が満足された場合に、主サイトDBS_rから各サイトDBS_iに対して更新情報を送信し、更新伝播を行うことにしている（図7参照）。実施の形態3においては、更新伝播を行うタイミングとして、主サイトDBS_rに設定された差異限定制約条件が満足された場合に加え、トランザクションが発生したサイトDBS側でコンテンツの内容に応じて更新伝播を行うタイミングを設定することができるようにする。このようにするのは、コンテンツによっては早期に公開すべきものや多少公開が遅れても良いとされるものがあり、例えば、台風情報等のように早期に公開すべき情報については5分以内に全てのサイトDBSに更新伝播を完了したいという要請に答えることができるようにするためである。

【0127】図10は、本発明の実施の形態3に係る分散型データベースシステムの一貫性管理方法を実現するための分散型データベースシステムの概念構成図である。図10に示す分散型データベースシステムは、図2に示した構成に加えて、更新伝播を行う際に、衛星1000を用いた放送を利用して主サイトDBS_rから各サイトDBS_iに更新情報を配信できるようにしたものである。一般に衛星1000等を利用して情報を配信することは例えばインターネットを利用する場合と比べてコストが高いが、緊急に全てのサイトDBSに更新伝播を行いたい場合等において放送により更新情報を配信することにより、全てのサイトDBSに対して更新情報を短時間にかつ一度に配信することができるという利点がある。

【0128】なお、図10において、1001a~1001eは、放送波を用いて更新情報を送信または受信するためのアンテナを、1002は実施の形態1および2で既に説明した各種情報を送受信するための通信回線をそれぞれ示している。

【0129】また、図10には、衛星1000を用いた放送により情報やり取りを行うことができる分散型データベースシステムを一例として示したが、これに加えて、または代えて地上波等、情報を送信可能な他の手段を用いることにしても良い。

【0130】次に、実施の形態3に係る分散型データベースシステムの一貫性管理方法について、サイトDBS

でレベルCの一貫性管理方法が設定されたトランザクションが発生した場合を例にとって説明する。なお、ここではレベルCの場合について説明するが、レベルBの場合にも同様に適用できることは明らかである。また、実施の形態1および2で既に説明した事項については、簡単に説明することにする。

【0131】図11は、更新伝播タイミング設定処理を示すフローチャートである。サイトDBS1のユーザは、複製データの更新を行う際に、例えば、予め用意された入力欄等に以下に説明する情報を設定する（S1101）。

【0132】ステップS1101においては、緊急度、公開指定日時、伝播遅延許容日時、伝播方式が設定されるものとする。緊急度には例えば以下のようなレベルがあり、これらのレベルは予めシステム構築者が設定する。したがって、ユーザは設定されている緊急度レベルの中から所望のレベルを選択し、緊急度の項目に設定することになる。

【0133】

- ・緊急度1：10分以内に更新伝播を行う。
- ・緊急度2：60分以内に更新伝播を行う。
- ・緊急度3：1日以内に更新伝播を行う。
- ・緊急度4：3日以内に更新伝播を行う。

【0134】公開指定日時の項目は、例えば報道ニュース等には明朝10時までは情報を公開してはならないという要求があることから設けられたものである。したがって、ユーザはこの項目に対して所望の日時を設定する。なお、この公開指定日時が指定された場合、主サイトDBS_rは、公開指定日時が経過した後更新された複製データに関する情報について更新伝播を行うようにする。また、これに代えて、公開指定日時の経過前に更新伝播を行うことにしても良い。いずれを用いるかについては、後述する更新伝播のスケジュールを設定する際（図13のステップS1304参照）に決定すれば良い。ただし、公開指定日時の経過前に更新伝播を行うことにする場合、各サイトDBSにおいて公開指定日時が経過する前に該当する更新後のデータを公開しないようにする手法を確立しておくことが必要となる。

【0135】伝播遅延許容日時の項目には、更新伝播の遅れが許される最大限の日時を設定する。

【0136】さらに、配信方式の項目には、通信（通信回線1002を用いる）、衛星放送、地上波放送等のいずれの方式を用いて更新伝播を行うかを設定する。

【0137】そして、ここではレベルCであるため、サイトDBS1においてトランザクションをコミットするかアボートするかが判定され、コミットの場合はサイトDBS_iの該当する複製データが更新される。なお、実施の形態2で説明した新規登録トランザクションの場合には、他のトランザクションとの競合のチェック処理は省略される。

【0138】コミットの場合、サイトDBS_iは、上述したように承認リクエストメッセージを主サイトDBS_rに通信回線1002を介して送信するが、その際、図12(a)および図12(b)に例として示すデータを承認リクエストメッセージ(図4に示したREQUESTメッセージと同一内容:レベルBの場合も同様)に添付する。その結果、図11のステップS1101で設定した緊急度等は、属性値として承認リクエストメッセージに設定されることになる。

【0139】主サイトDBS_rは、サイトDBS_iから承認リクエストメッセージを受信すると、他のサイトDBSで行われた更新処理と競合しないか否かを判定し、競合しないと判定した場合は以下の更新伝播管理処理を実行する。なお、新規登録トランザクションの場合は競合のチェック処理は省略される。また、競合のチェック処理と並行して更新伝播管理処理を実行することにしても良いが、他のサイトの更新処理と競合すると判定された場合は更新伝播管理処理を途中で終了することになる。

【0140】図13は、主サイトDBS_rによる更新伝播管理処理を示すフローチャートである。主サイトDBS_rは、サイトDBS_iから受信した承認リクエストメッセージから依頼元ID(サイトID)、依頼日時、更新OIDリスト、緊急度、公開指定日時、更新遅延許容日時および配信方式をコピーし(S1301)、さらに、依頼到着日時(主サイトDBS_rが承認リクエストメッセージを受信した日時)、データサイズ(送るべきデータのサイズ)を計測・設定し(S1302)、図14に示すような更新伝播管理情報を生成する。この更新伝播管理情報は、課金のためのデータとして利用することができる。

【0141】なお、ステップS1302において、承認リクエストメッセージに伝播遅延許容日時が設定されていない場合には、

依頼到着日時+指定された緊急度の許容時間(緊急度1なら10分)

という計算を行い、伝播遅延許容日時を設定する。

【0142】更新伝播管理情報を生成した後、主サイトDBS_rは、指定された緊急度が許容可能か否かについて、以下の基準に基づいて判定する(S1303)。

【0143】

・[依頼到着日時+指定された緊急度の許容時間]>伝播遅延許容日時の場合

・[公開指定日時+指定された緊急度の許容時間]>伝播遅延許容日時の場合

・(データサイズ×送付サイト数)の値が大きすぎる場合

なお、ここで、送付サイト数は、例えばサイトDBS_iが更新を行った複製データ(オブジェクト)と同一のオブジェクトに対する複製データを有する他のサイトDB

Sの数であり、どのサイトDBS_iが送付サイトに該当するかは主サイトDBS_rが判断する。

【0144】更新伝播管理情報中の値が上記3つの条件の少なくとも一つを満たすような場合、主サイトDBS_rは、指定された緊急度が許容可能ではないと判定し、許容不可能メッセージを生成してサイトDBS_i(依頼元のサイト)に送信し、図12に示したデータのみを再度送信するように要求する(S1307)。

【0145】そして、サイトDBS_i(依頼元のサイト)からメッセージに対する応答として図12に示したデータが送信され、そのデータを受信した場合(S1308)、主サイトDBS_rはステップS1301に戻って再度更新伝播管理処理を実行する。

【0146】一方、ステップS1303において、指定された緊急度が許容可能と判定した場合、主サイトDBS_rは、更新伝播管理情報中の更新遅延許容日時より前に更新情報を各サイトDBSに配信できるように、更新伝播を行うスケジュールを設定する(S1304)。例えば、主サイトDBS_rは、実施の形態1で説明した差異限定制約条件を用いて更新伝播を行うスケジュールを設定する。いずれの差異限定制約条件を用いるかについては、例えば、図14に示した更新伝播管理情報中の伝播遅延許容日時に更新伝播が間に合うことを第1の基準とし、これに加えて更新伝播のコストが低いものはどれかを第2の基準として判断する。そして、主サイトDBS_rは、選択した差異限定制約条件を用いて更新伝播を行うスケジュールリングを行い、伝播遅延許容日時を守れるか否かを判定し、守れないと判定した場合には、より早い時期に更新伝播を行うことができる差異限定制約条件を用いて再度更新伝播を行うスケジュールリングを行う。換言すれば、更新遅延伝播ルールとして各種の差異限定制約条件があるが、指定された緊急度を守ることができない可能性がある差異限定制約条件は採用しないことにする。

【0147】その後、主サイトDBS_rは、ステップS1304で設定したスケジュールに従い、更新伝播を実行する(S1305)。更新伝播は、更新伝播管理情報中に設定されている配信方式に従って更新情報を各サイトDBSに配信する。例えば、配信方式として「衛星」が設定されている場合は、衛星1000を用いて放送により更新情報を各サイトDBSに配信する。また、配信方式として「通信」が設定されている場合は、通信回線1002を介して更新情報を各サイトDBSに配信する。なお、通信回線1002を介して更新情報を各サイトDBSに配信する場合、承認リクエストメッセージを送信して来たサイトDBS(ここではサイトDBS_i)に対しては、通信コストを下げるために更新情報の配信は行わないようにする。

【0148】そして、主サイトDBS_rは、更新伝播の開始日時および完了日時のログと配信の遅れに関する情

報とを更新伝播管理情報中に設定する (S1306)。配信の遅れに関する情報とは、更新伝播管理情報中の伝播遅延許容日時からどれくらい遅れて配信されたかを示し、

伝播完了日時－伝播遅延許容日時

で計算される。計算した結果が正の場合は、伝播遅延許容日時から遅れて更新伝播が行われたことを示し、その値を更新伝播管理情報中の「配信の遅れ」に設定する。一方、計算した結果が負の場合は、伝播遅延許容日時の前に更新伝播が完了したことを示し、負の場合は配信の遅れは生じていないため、「0」を更新伝播管理情報中の「配信の遅れ」に設定する。

【0149】このように、実施の形態3に係る分散型データベースシステムの一貫性管理方法によれば、コンテンツの内容に適したタイミングで更新結果を各サイトDBSに配信することができるため、分散型データベースシステムの利便性の向上を図ることができる。

【0150】また、更新結果を各サイトDBSに配信する方法として、衛星1000を用いた放送による配信方法や通信回線1002を用いた通信による配信方法等を選択することができるため、緊急度に応じて適切な配信方法を選択でき、分散型データベースシステムの利便性の向上を図ることができる。

【0151】なお、詳細な説明については省略するが、レベルAの場合においても、放送により図4に示したコミットを配信することにしても良い。

【0152】以上説明した実施の形態1～3に係る分散型データベースシステムにおいては、トランザクションの種類毎に一貫性管理レベルについてレベルA～Cのいずれかを設定し、設定したレベルで一貫性管理を行うことにしたが、例えば、レベルCの一貫性管理レベルのみが利用可能なシステムやレベルAおよびレベルCの一貫性管理レベルが利用可能なシステム等、分散型データベースシステムをどのように運用するかによって、様々な設計・変更が可能なことは明らかである。

【0153】また、実施の形態1～3に係る分散型データベースシステムの一貫性管理方法は、図3～図5ならびに図11および図13に示したフローチャートの手順に従って、予め用意されたプログラムをコンピュータで実行することによって実現される。このプログラムは、ハードディスク、フロッピーディスク、CD-ROM、MO、DVD等のコンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行可能である。また、このプログラムは、上記記録媒体を介して、またはネットワークを介して配布することができる。

【0154】〔適用可能な応用分野〕以上説明した実施の形態1～3に係る分散型データベースシステムの一貫性管理方法は、以下に説明するような応用分野に適用することができる。ただし、本発明の分散型データベース

システムの一貫性管理方法を適用する応用分野を以下に説明するものに限定するものではない。

【0155】本発明の分散型データベースシステムの一貫性管理方法は、(1) マルチメディアのコンテンツを提供し、その視聴に対して課金を行うシステムであり、

(2) コンテンツの更新には時間的遅れを許し、(3) システム内に構築されるDB(データベース)としては、コンテンツDB・インデックス情報DB(2次情報)・利用者統計情報DB・課金情報DBという種類を有する分散型データベースシステムに適用することができる。

【0156】上記DBの特徴としては、以下のように考えられる。まず、コンテンツDBとしては、(1) データのサイズが超大で、データはマルチメディアのBLOB(Binary Large Object)であり、(2) 複製データが多数必要であり(コンテンツがマルチメディアであるため、サイズが大きく、アクセスコストを下げるためには複製データを多数生成することが必要)、(3) 読み出す情報は多少古くても良く、(4) 読み出し頻度は大であるものである。(5) 更新伝播は遅延しても良い、というものが考えられる。

【0157】インデックス情報DBとしては、(1) 複製データは多数必要であり、(2) 複製データデータのサイズは小～中であり、(3) 読み出す情報は多少古くても良く、(4) 読み出し頻度は大で、(5) 更新伝播は遅延しても良い、というものが考えられる。

【0158】利用者統計情報DBおよび課金情報DBとしては、(1) 複製データの数は少なくとも良く、

(2) データのサイズはシステムに依存し、(3) 読み出す情報は多少古くても良く、(4) 読み出し頻度は小であり、(5) 更新伝播は遅延しても良く、集計する側への伝播は遅延しても良い、というものが考えられる。ただし、利用者情報は紛失してもあまり問題とならないが、更新中に課金情報が紛失または欠落することは許されない。

【0159】上記分散型データベースシステムにおいて発生する典型的なトランザクションとしては、(1) 利用者によるコンテンツ検索(アクセス頻度大)、(2) コンテンツ提供者によるコンテンツの新規追加および修正(更新頻度は小)、(3) コンテンツ提供者によるインデックス情報の新規追加および修正(更新頻度は小)、(4) 課金情報の収集(各利用者に対する課金収集は頻度が小さいが、利用者数が膨大であるため、大きくなる)、および、(5) 利用統計情報の収集(各利用者に対する課金収集は頻度が小さいが、利用者数が膨大であるため、大きくなる)、が考えられる。

【0160】上記条件を満たすシステム例としては、

(1) 電子図書館システム(コンテンツは書籍、雑誌、CD-ROM、ビデオ等。インデックス情報は書誌情報。)、(2) ビデオ・オン・デマンド(VOD) シス

テム（コンテンツはビデオ。インデックス情報はビデオの各種属性。）、（3）カラオケシステム（コンテンツはMIDI等の音楽情報、動画情報および歌詞の文字情報等。インデックス情報は曲に関する各種属性。）、等が考えられる。

【0161】

【発明の効果】以上説明したように、本発明の分散型データベースシステムの一貫性管理方法（請求項1）によれば、予め複製データに対するリードまたは更新を含む各トランザクションの種類に、一貫性管理のレベルを表す第1レベル、第2レベルおよび第3レベルのいずれかを設定する一貫性レベル設定工程と、システム内においてトランザクションが発生した場合に、発生したトランザクションの種類に基づいて、一貫性管理のレベルを判定する判定工程と、判定工程で第1レベルであると判定された場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、第2のサイトから複数の第1のサイトへの更新伝播を即時行う第1の一貫性管理工程と、判定工程で第2レベルであると判定された場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、第2のサイトから複数の第1のサイトへの更新伝播を遅延させて行う第2の一貫性管理工程と、判定工程で第3レベルであると判定された場合に、弱い一貫性管理方法を用いて一貫性管理を行い、かつ、第2のサイトから複数の第1のサイトへの更新伝播を遅延させて行う第3の一貫性管理工程と、を含むことにより、トランザクションの種類に応じて最適なレベルの一貫性管理を行うことができるため、分散型データベースシステムの利便性の向上を図ることができる。

【0162】また、本発明の分散型データベースシステムの一貫性管理方法（請求項2）によれば、予め複製データに対するリードまたは更新を含む各トランザクションの種類に、一貫性管理のレベルを表す第1レベル、第2レベルおよび第3レベルのいずれかを設定する一貫性レベル設定工程と、システム内においてトランザクションが発生した場合に、発生したトランザクションの種類に基づいて、一貫性管理のレベルを判定する判定工程と、判定工程で第1レベルであると判定された場合に、read-onlyのトランザクションまたは更新を含むトランザクションのいずれであっても、厳格な一貫性管理方法を用いて一貫性管理を行い、第2のサイトから複数の第1のサイトへの更新伝播を即時行う第1の一貫性管理工程と、判定工程で第2レベルであると判定された場合に、read-onlyのトランザクションは緩和された一貫性管理方法を用いて一貫性管理を行い、更新を含むトランザクションは厳格な一貫性管理方法を用いて一貫性管理を行い、第2のサイトから複数の第1のサイトへの更新伝播を遅延させて行う第2の一貫性管理工程と、判定工程で第3レベルであると判定された場合に、read-onlyのトランザクションまたは更新

を含むトランザクションのいずれであっても、弱い一貫性管理方法を用いて一貫性管理を行い、第2のサイトから複数の第1のサイトへの更新伝播を遅延させて行う第3の一貫性管理工程と、を含むことにより、トランザクションの種類に応じて最適なレベルの一貫性管理を行うことができるため、分散型データベースシステムの利便性の向上を図ることができる。

【0163】また、本発明の分散型データベースシステムの一貫性管理方法（請求項3）によれば、請求項1または2に記載の分散型データベースシステムの一貫性管理方法において、第2および第3の一貫性管理工程における更新伝播の遅延について、第2のサイトにおいて予め設定されている更新伝播条件が満足された場合に更新伝播を行うこととしたことにより、更新が起こる毎に更新伝播を行う必要をなくすることができるため、データの一貫性保持のコストを削減することができる。

【0164】また、本発明の分散型データベースシステムの一貫性管理方法（請求項4）によれば、請求項1～3のいずれか一つに記載の分散型データベースシステムの一貫性管理方法において、第2および/または第3の一貫性管理工程における更新伝播の遅延とは、第1のサイトにおいてトランザクション毎に更新伝播を行うタイミングを定めた更新伝播条件が設定され、第2のサイトにおいて更新伝播条件を満足するように複数の第1のサイトへの更新伝播を行うこととしたため、ユーザの所望の更新伝播条件を設定することができる。したがって、ユーザの所望の更新伝播条件を設定することができるため、更新伝播がいつ行われるかを容易に予測することができ、分散型データベースシステムの利便性のさらなる向上を図ることができる。

【0165】また、本発明の分散型データベースシステムの一貫性管理方法（請求項5）によれば、請求項4に記載の分散型データベースシステムの一貫性管理方法において、第2および/または第3の一貫性管理工程が、少なくとも通信回線および放送波を用いて第2のサイトから複数の第1のサイトへの更新伝播を行うことができ、更新伝播条件が、通信回線および放送波のいずれを用いて更新伝播を行うかについての指定を含むことにより、更新伝播条件に応じて適切な方法で更新伝播を行うことができるため、分散型データベースシステムの利便性の向上を図ることができる。すなわち、緊急に更新伝播を行う必要がある場合にはコストが高くても放送波を用いて更新伝播を行うことにし、時間的に余裕がある場合にはコストを抑えて通信回線を用いて更新伝播を行うという選択を行うことができる。

【0166】また、本発明の分散型データベースシステムの一貫性管理方法（請求項6）によれば、請求項1～5のいずれか一つに記載の分散型データベースシステムの一貫性管理方法において、判定工程が、一貫性管理のレベルを判定することに加えて、データの新規登録可否

かを判定し、新規登録であると判定した場合に、設定されているレベルに関係なく、第3レベルと判定するため、データを新規登録する際の処理効率を向上させることができる。

【0167】さらに、本発明のコンピュータ読み取り可能な記録媒体（請求項7）によれば、請求項1～6のいずれか一つに記載の分散型データベースシステムの一貫性管理方法の各工程をコンピュータに実行させるためのプログラムを記録したため、このプログラムをコンピュータに実行させることにより、分散型データベースシステムの一貫性管理のレベルを複数段階に分け、トランザクションの種類に応じて異なる一貫性管理を行うことができ、分散型データベースシステムの利便性の向上を図ることができる。

【図面の簡単な説明】

【図1】集中管理または分散管理と、強い一貫性管理または弱い一貫性管理との組み合わせによって、分散型データベースシステムを構築した場合の長所および短所が変化することを示した説明図である。

【図2】本発明の実施の形態1に係る分散型データベースシステムの一貫性管理方法を実現するための分散型データベースシステムの概念構成図である。

【図3】本発明の実施の形態1に係る分散型データベースシステムの一貫性管理方法において、（a）は一貫性管理のレベルを設定する手順を、（b）は（a）に示すようにして設定した一貫性管理のレベルに基づいて一貫性管理方法を実行する手順を示すフローチャートである。

【図4】本発明の実施の形態1に係る分散型データベースシステムの一貫性管理方法において、各サイトから主サイトに送信されるREQUESTメッセージの説明図である。

【図5】本発明の実施の形態1に係る分散型データベースシステムの一貫性管理方法において、主サイトから各サイトに送信されるコミットメッセージの説明図である。

【図6】本発明の実施の形態1に係る分散型データベースシステムの一貫性管理方法において、主サイトから各サイトに送信されるアポートメッセージの説明図である。

【図4】

REQUEST
・サイトID
・トランザクションID
・（操作、対象オブジェクトID、t（O _i 、T） 【更新後の値】）の3組または4組のリスト

【図7】本発明の実施の形態1に係る分散型データベースシステムの一貫性管理方法において、（a）は差異限定制約条件を設定する手順を、（b）は（a）に示すようにして設定した差異限定制約条件に基づいて更新伝播を行う手順を示すフローチャートである。

【図8】本発明の実施の形態1に係る分散型データベースシステムの一貫性管理方法において、一貫性管理のレベルA～Cの特徴をまとめた説明図である。

【図9】本発明の実施の形態2に係る分散型データベースシステムの一貫性管理方法において、（a）は一貫性管理のレベルおよび新規登録トランザクションを設定する手順を、（b）は（a）に示すようにして設定した一貫性管理のレベルおよび新規登録トランザクションの設定に基づいて一貫性管理方法および新規登録を実行する手順を示すフローチャートである。

【図10】本発明の実施の形態3に係る分散型データベースシステムの一貫性管理方法を実現するための分散型データベースシステムの概念構成図である。

【図11】本発明の実施の形態3に係る分散型データベースシステムの一貫性管理方法において、更新伝播タイミング設定処理を示すフローチャートである。

【図12】本発明の実施の形態3に係る分散型データベースシステムの一貫性管理方法において、更新伝播タイミングをサイト側で設定する場合に、リクエストメッセージに添付されるデータを示す説明図である。

【図13】本発明の実施の形態3に係る分散型データベースシステムの一貫性管理方法において、主サイトによる更新伝播管理処理を示すフローチャートである。

【図14】本発明の実施の形態3に係る分散型データベースシステムの一貫性管理方法において、主サイトによって管理される更新伝播管理情報の一例を示す説明図である。

【符号の説明】

DBS, 主サイト

DBS₁, DBS₂, DBS₃, ..., DBS_n

サイト

1000 衛星

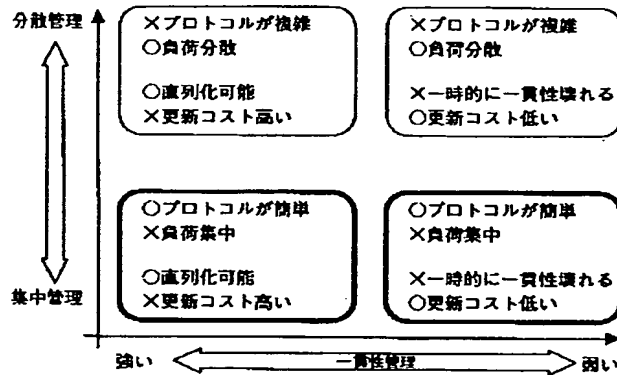
1001a~1001e アンテナ

1002 通信回線

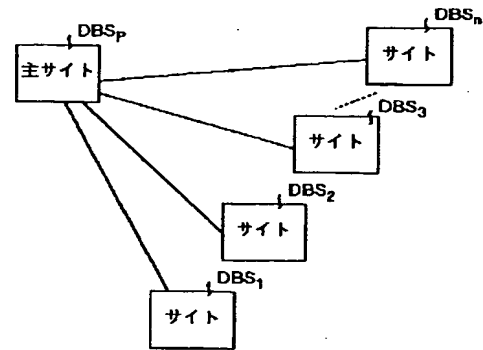
【図5】

COMMIT
・サイトID
・トランザクションID
・（対象オブジェクトID、t（O _i ）、更新後の値） の3組のリスト

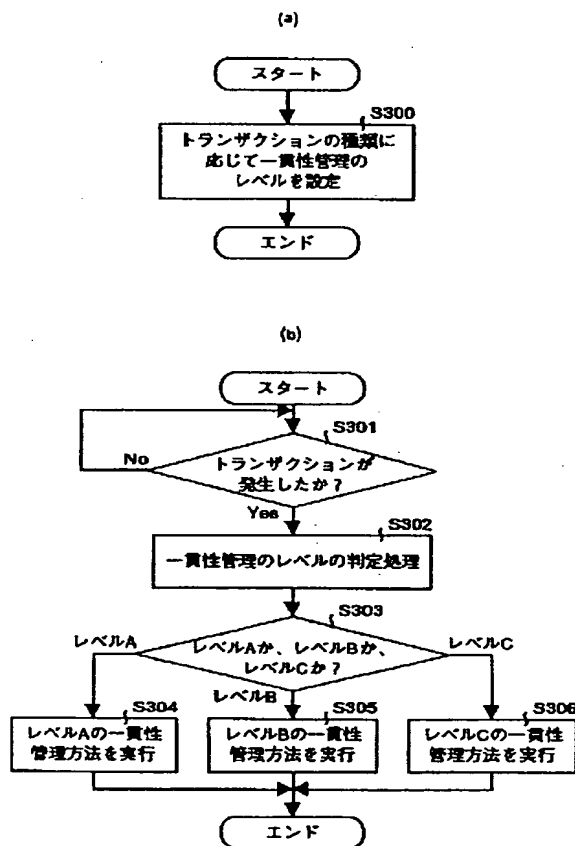
【図 1】



【図 2】



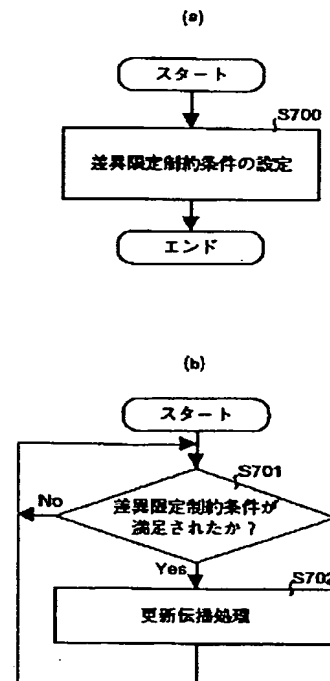
【図 3】



【図 6】

ABORT
・ サイトID
・ トランザクションID
・ (対象オブジェクトID、t (O _o)、現在の値O _o) の3組のリスト

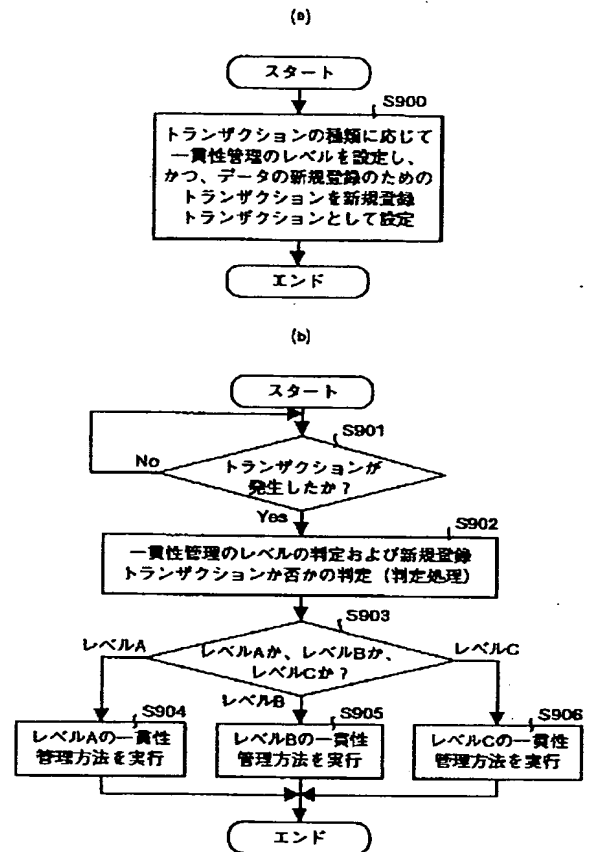
【図 7】



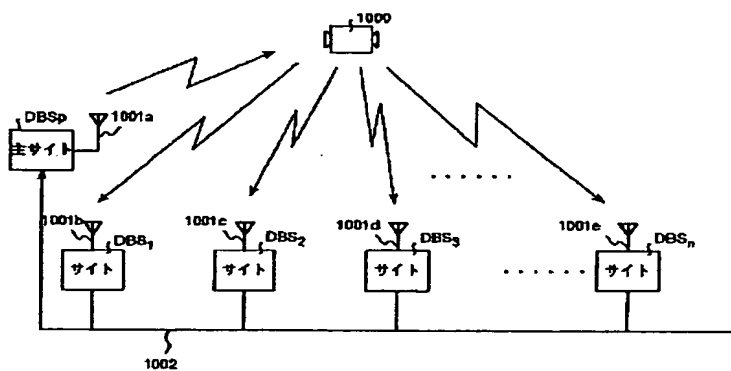
【図 8】

レベル	read-onlyの トランザクション	更新を含む トランザクション	主サイトからの 更新伝播の遅延
A	厳格	厳格	即時
B	緩和 1コピー直列化可能	厳格	遅延させる (差異限定制約)
C	緩和 1コピー直列化可能	緩和	遅延させる (差異限定制約)

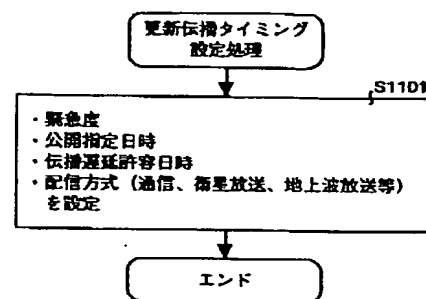
【図 9】



【図 10】



【図 11】



【図 1 2】

(a)

依頼元ID 8945	送付先 (主サイト) ID 56	依頼日時 1998/3/10 20:18:10	更新OIDリス上 45,682,589,439,902	緊急度 1	公開指定日時 指定無し	伝播遅延許容日時 1998/3/10 20:30:00	配信方式 衛星
---------------	---------------------	----------------------------	--------------------------------	----------	----------------	--------------------------------	------------

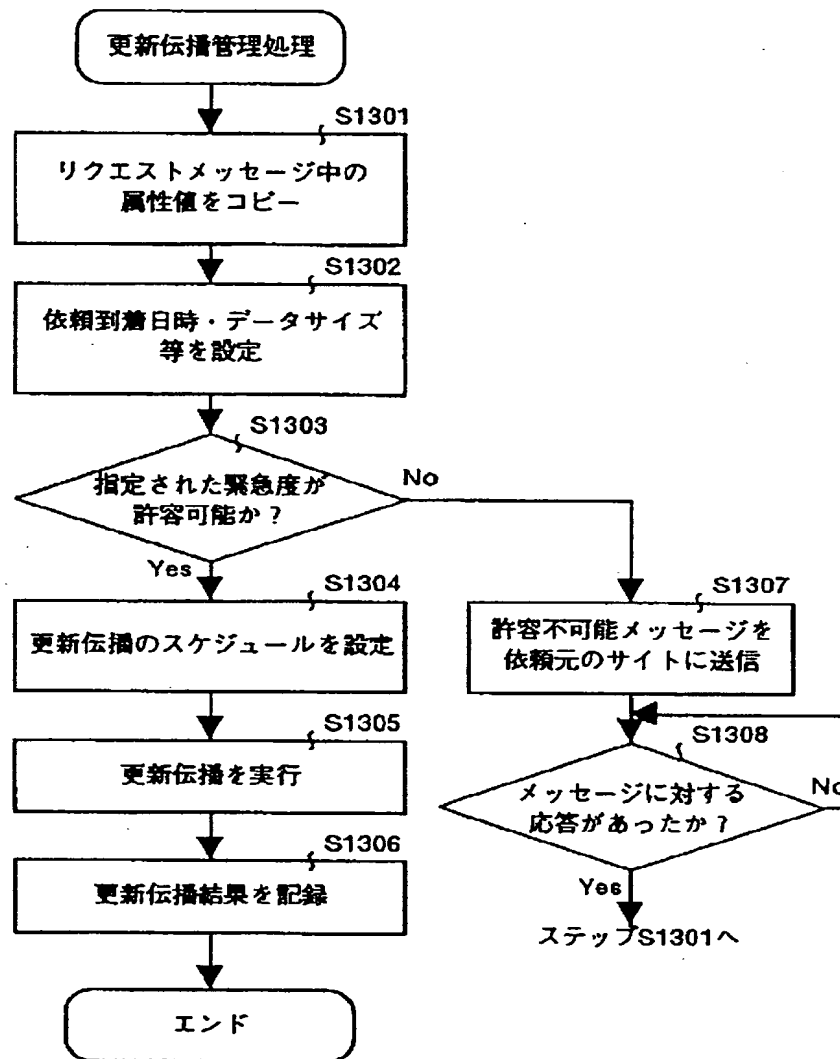
(b)

依頼元ID 9824	送付先 (主サイト) ID 90	依頼日時 1998/3/10 20:23:15	更新OIDリス上 882,678	緊急度 1	公開指定日時 1998/4/1 10:00:00	伝播遅延許容日時 1998/4/1 10:10:00	配信方式 通信
---------------	---------------------	----------------------------	---------------------	----------	-----------------------------	-------------------------------	------------

【図 1 4】

依頼元ID 8945	依頼到着日時 1998/3/10 20:18:13	依頼日時 1998/3/10 20:18:10	更新OIDリス上 45,682,589,439,902	緊急度 1	公開指定日時 指定無し	伝播遅延許容日時 1998/3/10 20:30:00	データサイズ 310MB
送付サイト数 120	配信方式 衛星	伝播開始日時 1998/3/10 20:25:21	伝播完了日時 1998/3/10 20:25:21	配信の遅れ 0			

【図 1 3】



フロントページの続き

(51) Int. Cl.⁶

識別記号

F I

G 0 6 F 15/401

3 4 0 A

(72) 発明者 前田 薫
東京都大田区中馬込 1 丁目 3 番 6 号 株式
会社リコー内

(72) 発明者 池田 哲也
東京都大田区中馬込 1 丁目 3 番 6 号 株式
会社リコー内